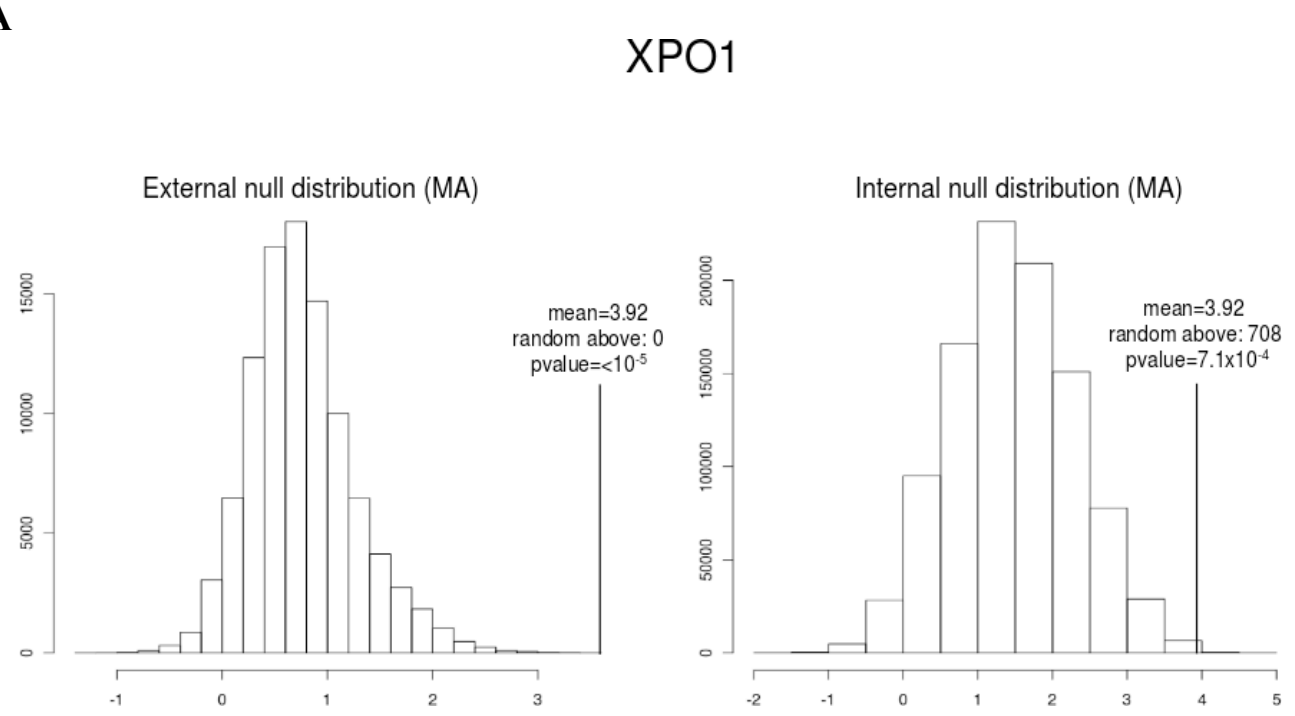
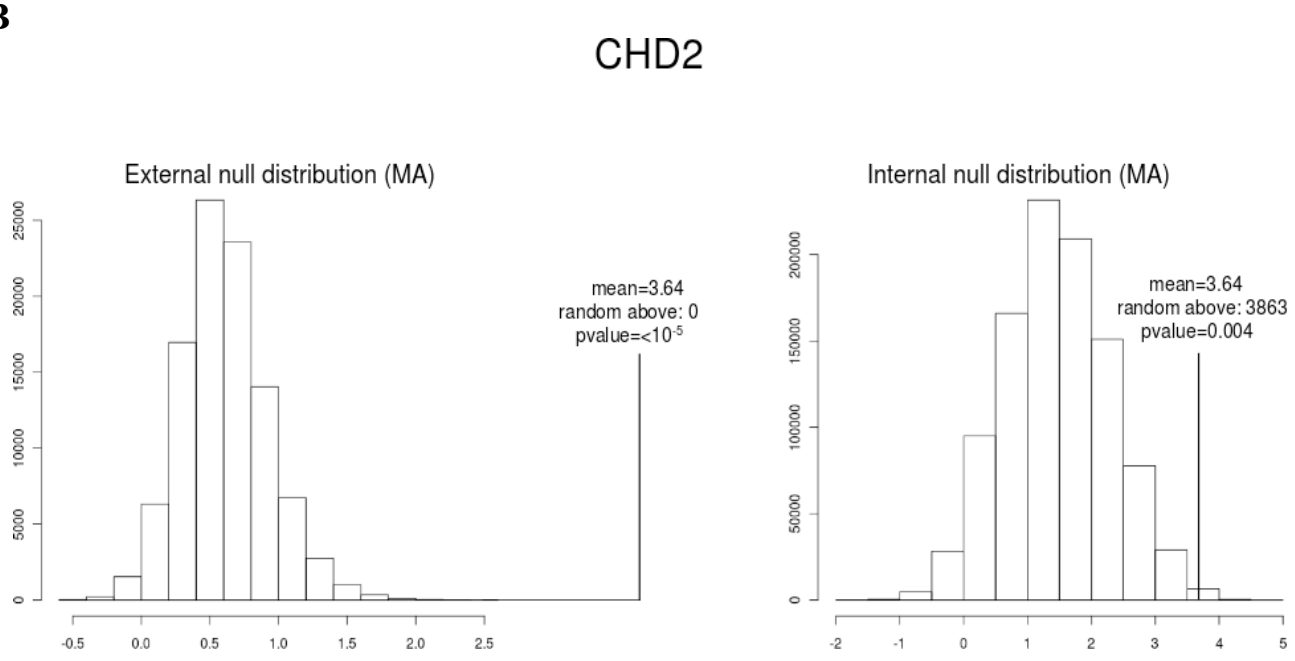
Supplementary Figure S1

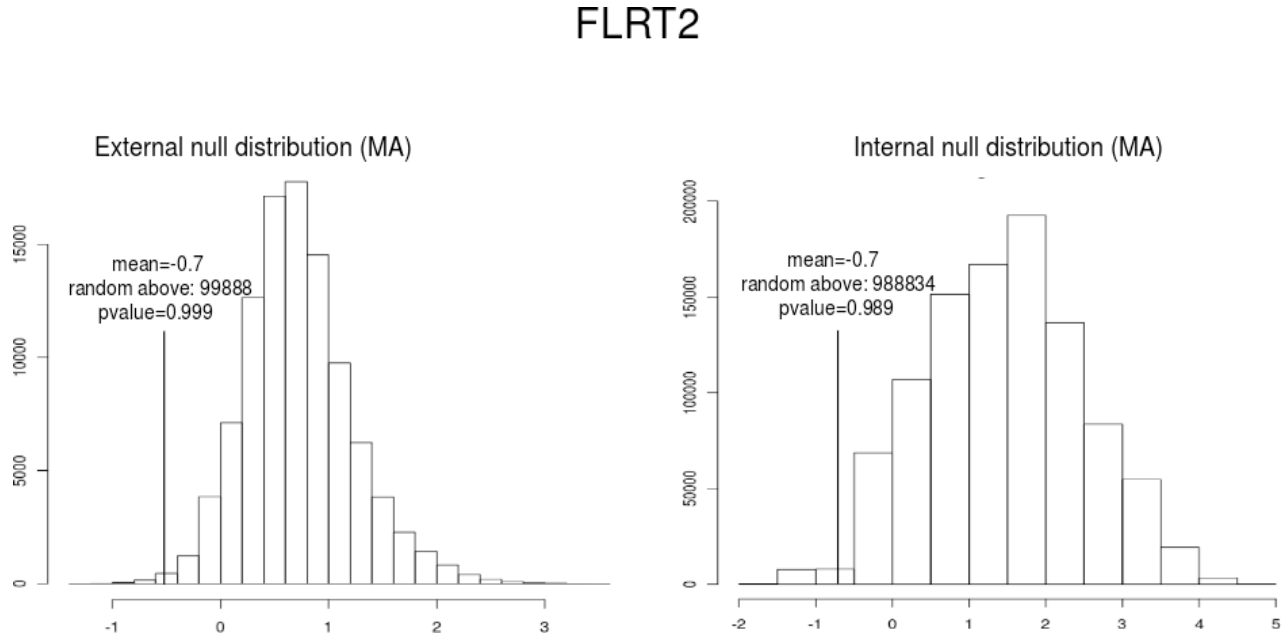
A



B

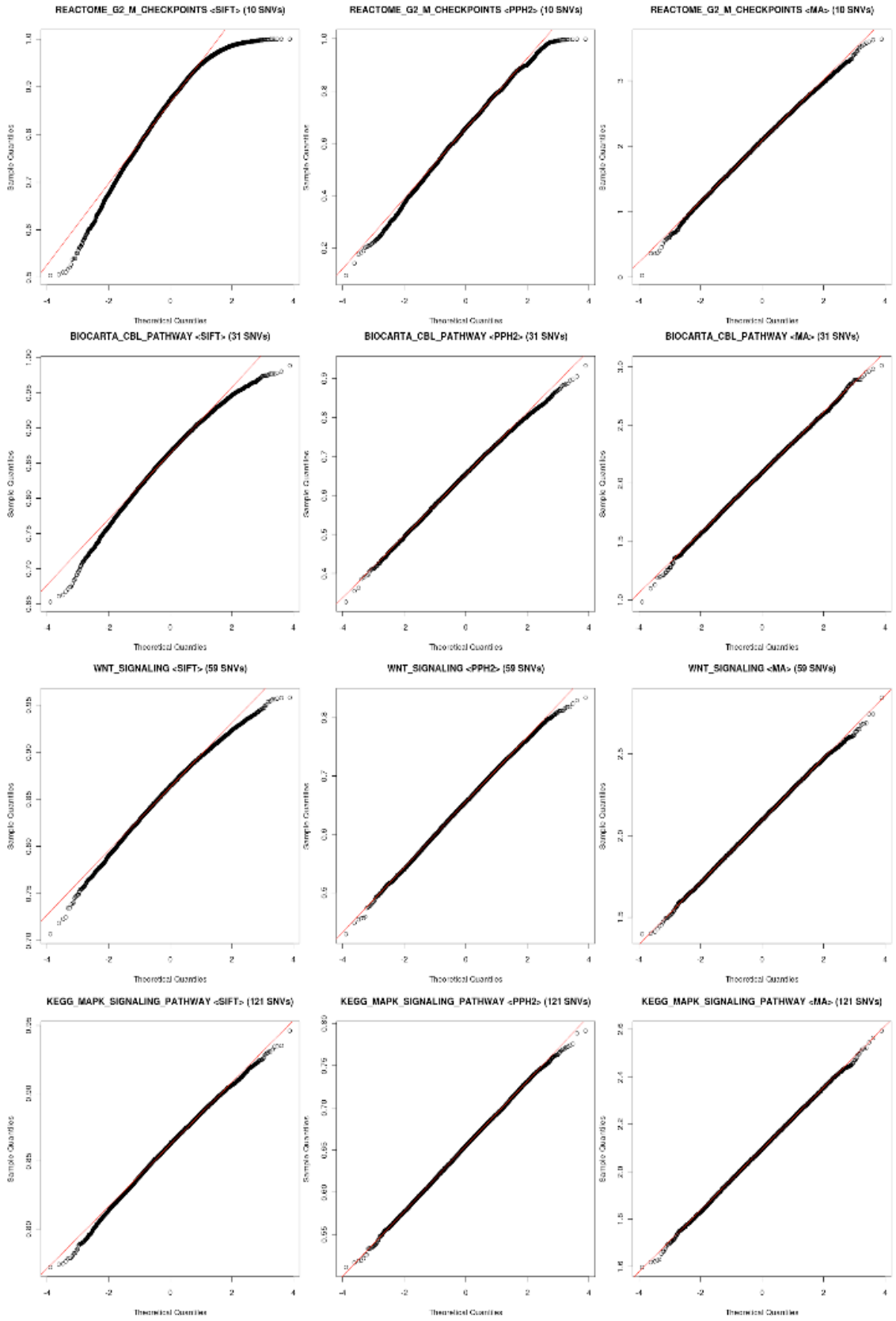


C



Supplementary Figure S1. Comparing the assessment of the FM bias pvalue with the external and internal null distributions in three genes from the cll dataset. **A**, XPO1; **B**, CHD2; **C**, FLRT2. In each panel, the **left histogram** corresponds to the external null distribution, namely 100000 means of the MA scores of groups of mutations of the same size observed for the gene in question randomly sampled from the SNVs found in genes with the same slimGO across human populations by the 1000genomes project. On the other hand, the **right histogram** corresponds to the internal null distribution, namely one million means of the MA scores of groups of mutations of the same size observed for the gene in question across all cll samples. For each gene, the panels show the mean MA score of the mutations it receives in cll, the number of random means in the null distribution (internal or external) above that observed mean, and the empirical pvalue hence computed.

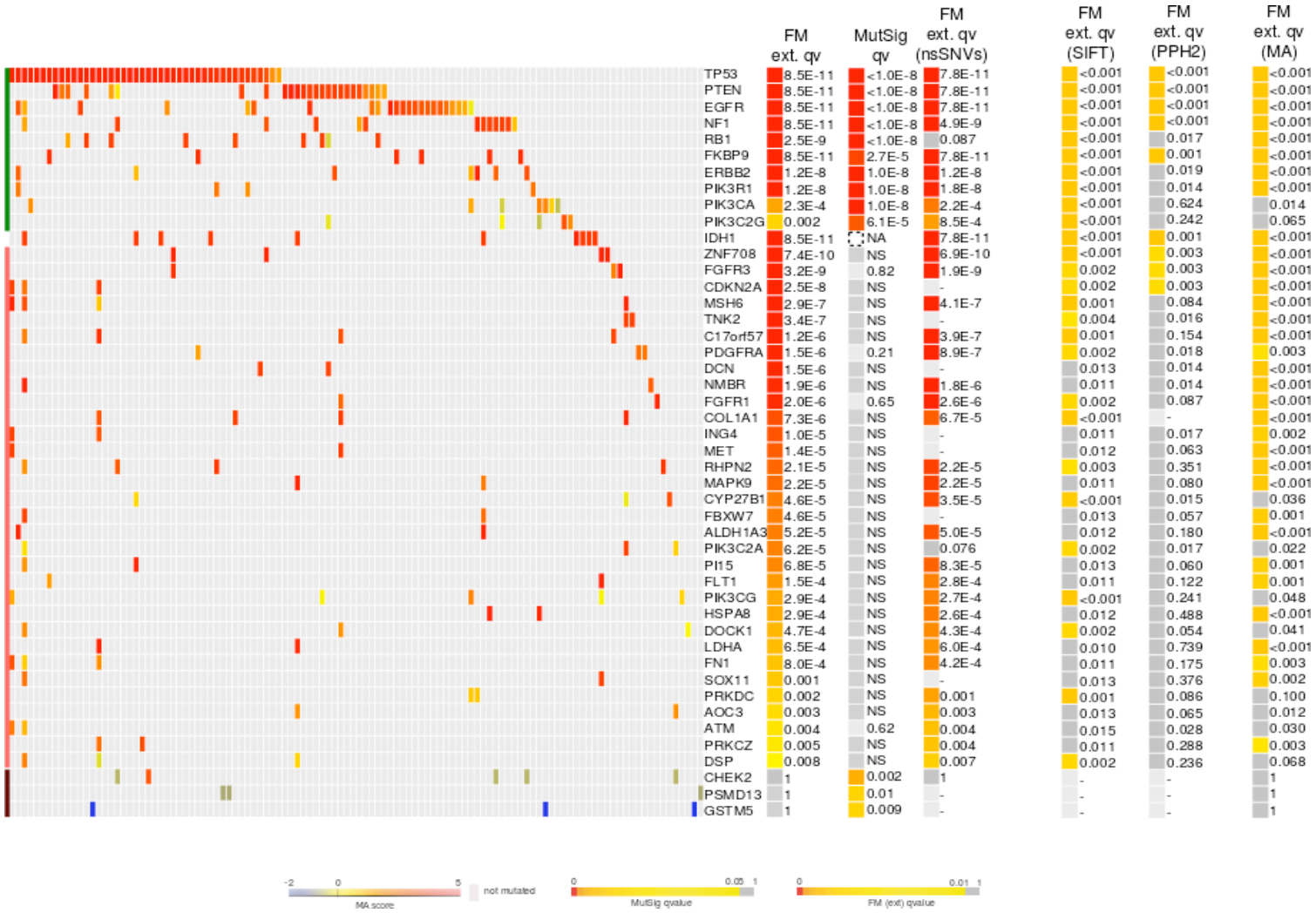
Supplementary Figure S2



Supplementary Figure 2. QQplots of random null distributions of somatic mutations produced for several gbm pathways. Each distribution contains 10000 means obtained from groups of mutations of the same size as the analyzed pathway and randomly sampled from all mutations identified across all tumors. (See main text.) Each row illustrates the null distributions produced to evaluate the FM bias of a given pathway; each column corresponds to one of the FI scores employed to compute the FM Zscores of pathways. (From left to right: SIFT, PolyPhen2, MutationAssessor.)

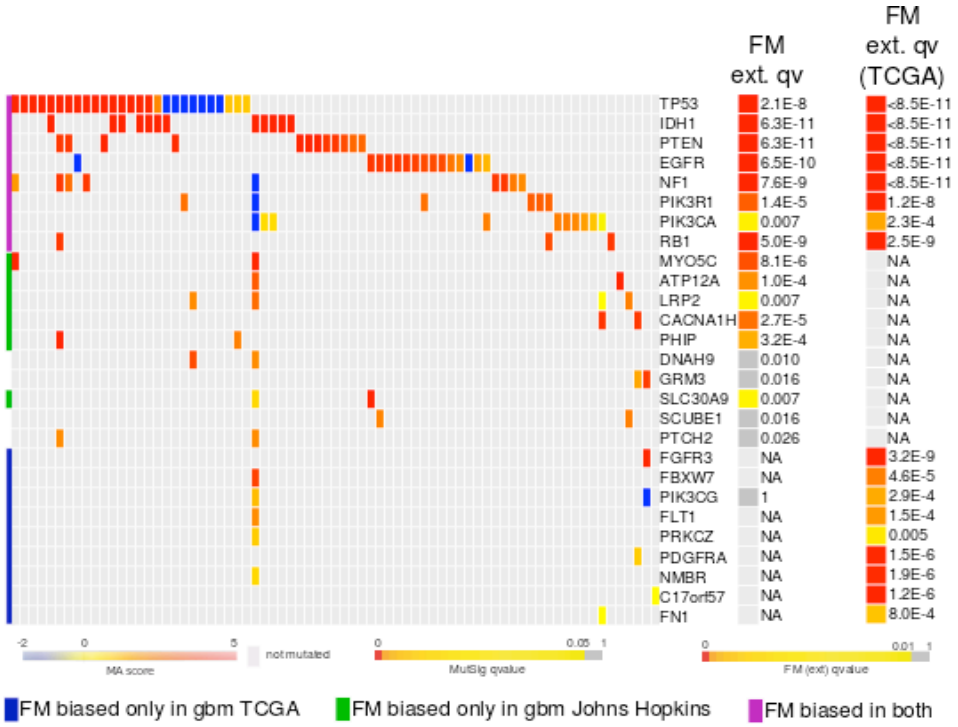
For instance, the top right qqplot illustrates the distribution of 10000 means obtained from the MutationAssessor scores of groups of 10 somatic mutations (the number found in the G2/M checkpoints pathway from Reactome).

Supplementary Figure S3



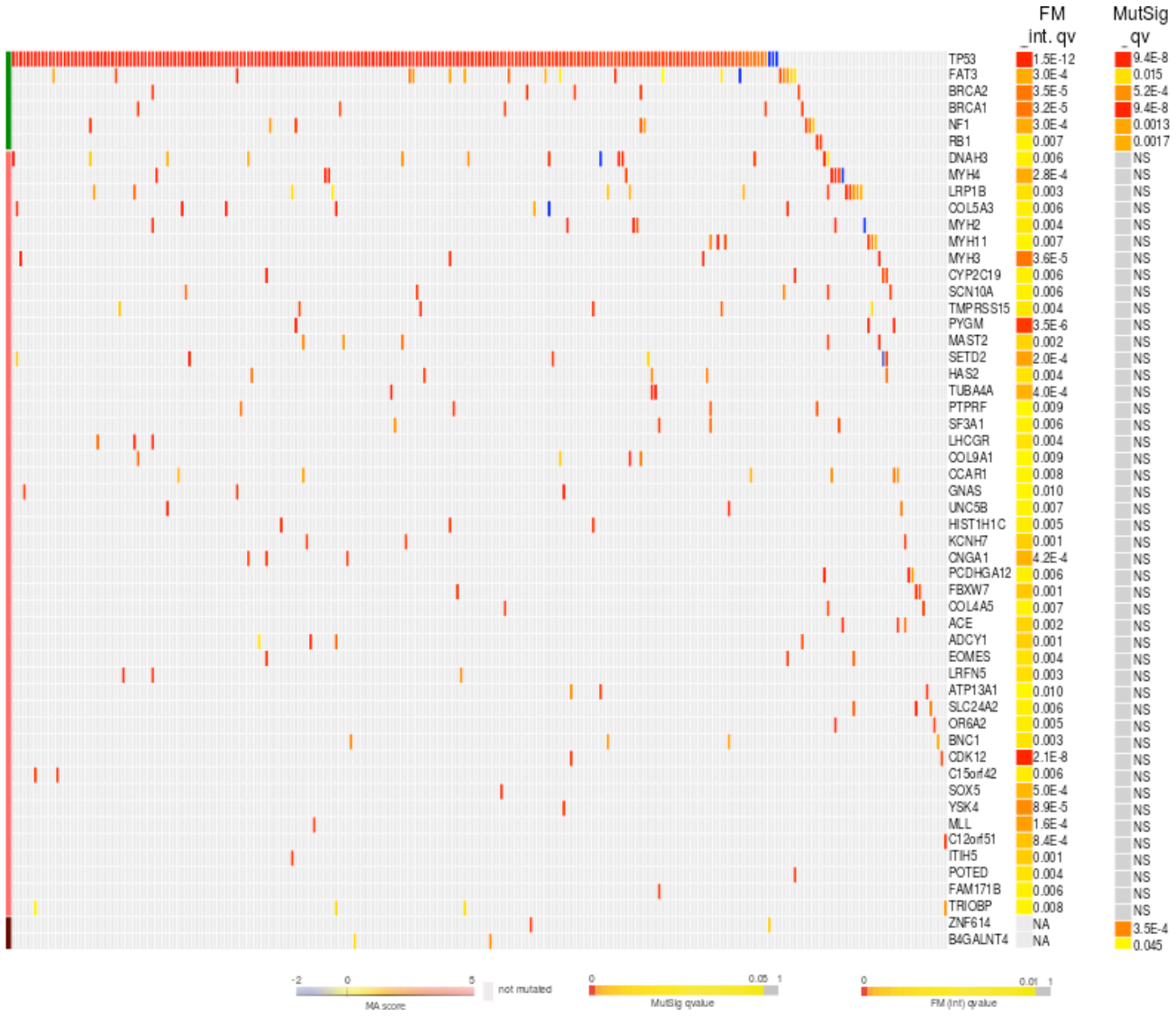
Supplementary Figure S3. Complete list of FM biased genes detected by Oncodrive-fm in 135 glioblastoma multiforme samples from TCGA. **FM ext. qv**, corrected pvalues of the FM bias analysis using the external null distribution. **MutSig qv**, corrected pvalues of the mutation recurrence analysis (implemented by MutSig). **FM ext. qv nsSNVs**, corrected pvalues of the FM bias analysis using only nsSNVs to compute the average FI and the external null distribution. **FM ext. qv (SIFT)**, corrected pvalues of the FM bias analysis using only SIFT scores. **FM ext. qv (PPH2)**, corrected pvalues of the FM bias analysis using only PolyPhen2 scores. **FM ext. qv (MA)**, corrected pvalues of the FM bias analysis using only MutationAssessor.

Supplementary Figure S4



Supplementary Figure S4. FM biased genes detected by Oncodrive-fm in 77 glioblastoma multiforme samples from Johns Hopkins University (ref.). **FM ext. qv**, corrected pvalues of the FM bias analysis using the external null distribution. **FM ext. qv (TCGA)**, corrected pvalues of the FM bias analysis of the TCGA gbm dataset employed in the paper to validate the Oncodrive-fm approach (Figure 1A of the main text). Genes found to be FM biased in both datasets are marked purple at the extreme right of the figure, while those FM biased only in one are marked either blue (gbm TCGA) or green (gbm Johns Hopkins).

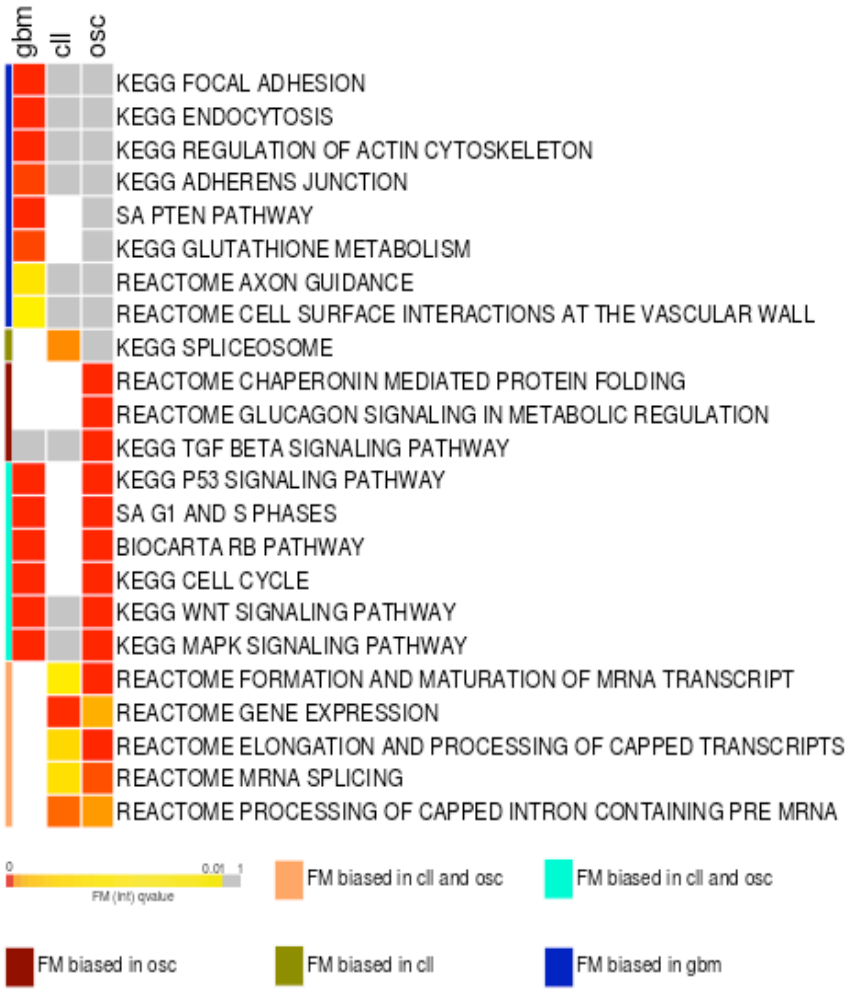
Supplementary Figure S5



Supplementary Figure S5. Significantly FM biased genes detected by Oncodrive-fm in 316 ovarian serous carcinoma samples from TCGA (complete list). As in Figure 2A of the main paper, genes are classified as RFM (green bar at the extreme left of the panel), IRFM (pink) or RnFM (dark red).

FM int. qv, corrected pvalues of the FM bias analysis using the internal null distribution. **MutSig qv**, corrected pvalues of the mutation recurrence analysis (implemented by MutSig). All heatmaps were built using Gitools (ref.) and include only genes with at least three mutated samples (two in ZNF614 and B4GALNT4).

Supplementary Figure S6



Supplementary Figure S6. Examples of significantly FM biased pathways in gbm, cll and osc.